**Simulating Natural Viewing Conditions for 3D Perception in Lightness Discrimination Experiments Using VR**

Devin Reynolds

Dr. Vijay Singh

May 3, 2023

North Carolina Agricultural and State Technical University

National Science Foundation (BCS – 2054900)

**Abstract**

The human visual system provides a relatively constant representation of the color of an object despite changes in properties that are irrelevant to the object. Understanding and quantifying the visual system’s ability to retain this constancy despite variation in object irrelevant properties is a goal of vision science. In previous works, color constancy experimentation involved presenting visual stimuli on a monitor. This method of presenting visual stimuli is not representative of natural viewing conditions. In natural vision, humans use two frontal eyes to obtain two different images of the same scene from slightly different angles. The natural disparity between these two images allows animals to accurately judge the distance of objects. When displaying on a monitor, the same visual stimuli is presented as a single 2D image to both eyes. These 2D image lack the perception of the third dimension. To simulate natural vision, sets of two images, stereoscopic pairs, are rendered of the same scene from two slightly different positions. The disparity of these two images is representative of the natural disparity between the two eyes in vision. The stereoscopic pairs can then be displayed to their corresponding eyes to give an accurate perception of distance within the rendered image. Virtual Reality (VR) can be used control what is displayed to each eye. Game engines like Unity are a popular choice when working with VR. Unity provides a framework that can develop psychophysical experiments and display stereoscopic pairs. Unity provides a 3D scene where rendered objects can be displayed. The stereoscopic pairs can be placed inside the Unity scene as 2D planes alongside two cameras corresponding to the left and right eyes of the VR headset. The rendering of the cameras must then be limited to only their corresponding image. Viewing the scene through a VR headset should now provide accurate information of the third dimension in the images. Now that a system for viewing stereoscopic images has been made, color or lightness constancy can be studied under natural viewing conditions. Two test conditions that have been previously studied on a monitor are ran using this VR set-up: no variation and natural variation. For no variation, the background objects in the images have a fixed color. For natural variation, background objects are allowed to change in their color. Similar to previous experiments, it was found that subject performance decreased with an increase in variation.

**Introduction**

The human visual system provides a relatively constant representation of the color of an object despite changes in properties that are irrelevant to the object (Brainard, 2009). However, this constancy is not always preserved (Brainard, 2009). Understanding and quantifying the visual system’s ability to retain this constancy despite variation in object irrelevant properties is a goal of vision science. In previous works, human observers viewed rendered images of a monochromatic sphere within a three-dimensional (3D) scene containing naturalistic background objects and reported the image that contained the lighter sphere. These images were displayed to observers as two-dimensional (2D) images on a color calibrated monitor. This, however, is not representative of natural human vision. In this work, virtual reality (VR) and Unity are used to view the rendered images under more naturalistic conditions.

In natural vision, humans perceive the world in three dimensions, of which, distance from the observer is not present in the 2D images displayed on monitors. The images taken by human eyes remain 2D yet humans can still perceive the third dimension (McCoun and Reeves, 2010). This third dimension information is normally provided to humans from a combination of monocular (one eye) and binocular (two eye) depth cues. Here, two binocular cues will be exploited to provide distance information to rendered 2D images which are stereopsis and convergence.

Humans use two frontal eyes to obtain two different images of the same scene from slightly different angles. In stereopsis, the natural disparity between these two images allows humans to accurately triangulate the distance of objects (McCoun and Reeves, 2010). This disparity is inversely dependent on the distance of an object from the observer (McCoun and Reeves, 2010). As an object is moved further away from an observer, the disparity decreases, and as an object is moved closer to an observer, the disparity increases. During stereopsis, the eyes focus on the same point, and in doing so, will stretch the extrinsic muscles of the human eye when viewing objects at close distances (<10m) (Cassin and Solomon, 1990). This is called convergence. The sensations of stretching the extrinsic muscles during convergence provide distance information. Stereopsis of far objects provides distance information of objects relative to one another and gives an illusion of distance. Convergence provides absolute distance information of an object from an observer.

When displaying images on a monitor, the same visual stimuli is presented as a single 2D image to both eyes. From the binocular cues, a second image is needed to provide distance information to an observer. To simulate natural vision, a second image, of the same scene, can be rendered at a slightly different position (63mm apart) to represent the distance between the two human eyes. The disparity of these two images is representative of the natural disparity between the two eyes in vision. These two 2D images form a pair called stereoscopic pairs. These pairs represent what the human eyes see. The stereoscopic pairs can be displayed to their corresponding eyes to give an accurate perception of distance within the rendered scene.

**Image Rendering**

The images are generated using a software called Virtual World Color Constancy (VWCC) (https://github.com/BrainardLab/VirtualWorldColorConstancy). VWCC is written using Matlab and uses the Mitsuba renderer to render images. Before rendering the image, the scene must be created. The 3D geometry and light sources can be inserted in user specified locations using the software, Blender. As in previous works, the shape and position of the 3D geometry and light sources are fixed.

The reflectance spectra of the objects are generated using random sampling from a statistical model of surface reflectances, derived from datasets of natural world objects, as described in Singh et. al. (2018). The natural datasets are approximated using principal component analysis (PCA). The dataset is then projected along the PCA directions with the largest 6 eigenvalues. The resulting distribution is then approximated as a multivariate normal distribution as described in Singh et. al. (2018). The reflectance spectra for the background objects are randomly sampled from this distribution.

For the spectra of the light source, CIE standard illuminant D65, which corresponds to the average midday light in western Europe, is used. Standard illuminant D65 is normalized by its mean to get the shape of the spectra. The normalized D65 spectra is used for the spectra of the light sources.

The object reflectance spectra and the light source illumination spectra, acquired from above, are applied to the objects and the light source in the scene. Then a stereoscopic pair of 2D multispectral images of the scene are rendered using the Mitsuba renderer at 31 wavelengths linearly spaced between 400 nm and 700 nm. These two images are rendered at two camera positions that are 63mm apart to represent the average interpupillary distance (Figure 1).

A sample image library is made of 10 stereoscopic pairs per 11 lightness values of the spherical target object between 0.35 and 0.45, for a total of 110 stereoscopic pairs (220 total images). To present these multispectral stereoscopic images, they are first converted to PNG images. The PNG images can then be displayed using virtual reality.



(Figure 1) An example image of a rendered stereoscopic pair from the larger image library. The two images correspond to the two slightly different images that the eyes naturally see. The second image is rendered 63mm to the right as 63mm is the average distance between the eyes. The geometry and light sources can be placed at user specified locations. The spectra of the objects and light sources can also be controlled. A spherical tar

**Virtual Reality**

Virtual Reality (VR) headsets are devices worn on your head that provide an immersive experience of a simulated 3D virtual environment. VR displays 3D environments using a combination of hardware and software. The hardware component of a VR headset consists of two small screens and lenses placed in front of each eye, which create a stereoscopic image. These 3D environments are rendered twice, once for each screen, at slightly different positions to produce stereoscopic vision.

VR headsets fit into one of two categories, either standalone VR headsets which are made with a built-in graphics possessing unit and processor or more powerful headsets that need to be connected to a computer. The Meta Quest 2, which is used in this work, was designed to be a standalone VR headset, however it can still be connected to a computer. Connecting the Quest 2 to a computer gives greater control over what can be displayed to the headset.

Since the interest is running psychophysical experiments using VR and stereoscopic images, a program must be designed to run the experiments and display the images. To make a program for VR, first a development platform needs to be chosen. There are several platforms available for developing VR apps, like game engines such as Unity or Unreal Engine.

**Unity**

Unity is a software that is used to make interactive experiences, of which, the most common are video games. Using unity, virtual worlds can be created that can then be displayed to a VR headset. Unity provides a framework that can develop psychophysical experiments and display stereoscopic pairs for VR.

To use Unity with VR, two cameras corresponding to the left and right displays of the headset are needed. The method of creating the cameras for the headsets are different depending on which VR headset is used. For the Meta Quest 2, the Oculus Integration package should be installed from the asset store (https://assetstore.unity.com/packages/tools/integration/oculus-integration-82022). The Oculus Integration package contains tools and assets, including a VR camera object, specifically designed for use with Meta’s VR devices. The camera, called OVRCameraRig, can be placed within a 3D scene along side two 2D planes for the 2D stereoscopic images that were rendered. These planes are blank square objects that can be used to display images to the camera. The planes that are placed inside of the scene need to be the only thing that is rendered by the camera. This can be done using some built-in Unity tools.

The Layers tool assigns a layer, which is a label, to each object in the scene. It can then be chosen how these objects, in a specific layer, interact with objects in other layers. Another unity tool, the culling mask is used to control which objects in a scene are visible to a camera. The culling mask, used in tandem with the layers tool, can render selective parts the scene by layers. These tools can be used to display the images in a stereoscopic pair to their corresponding cameras, i.e., the left image is displayed only to the left camera and the right camera is only displayed to the right camera.

Images must first be imported to unity before use, however, by default imported images are compressed. Compressing the images changes how the images appear, i.e. makes them blurry, which is detrimental to a color based experiment. The import compression can be disabled in Unity’s settings. Afterwards, the images can be moved directly into the Unity project directory and the PNG images will be automatically imported.

The 2D planes placed in the scene are still blank objects. Before the images can be placed in the scene, a material needs to be created for the images. A material is a file that contains the information about how an object is rendered. Materials determine the object's appearance by specifying its color, texture, transparency, and other visual properties. Two materials need to be created, one for the left eye image and one for the right eye image. These materials can then be applied to the two planes, which changes the appearance of the planes to the rendered images. Now that a method for displaying stereoscopic images has been created, the task can be defined.

**Lightness Discrimination Task**

Similar to the previous monitor experiment, two stereoscopic pairs are shown to the subject and they report on which image they think contains the lighter target object (Figure 2). Four images, two stereoscopic pairs, in total are displayed but the pairs appear as a single image when viewed through the headset. Of these two pairs, one is called a standard image and the other is called the comparison image. The target object of the standard image has a fixed lightness value of 0.40, while the target object of the comparison image has one of 11 linearly spaced lightness values between 0.35 and 0.45.

The subject reports on the lightness for 30 pairs of images for each of the 11 comparison image lightness values. Lightness discrimination thresholds are then chosen to be the 50%-76% difference of the comparison image lightness along a cumulative gaussian fitted to the proportion of times the comparison image was chosen (Figure 3).

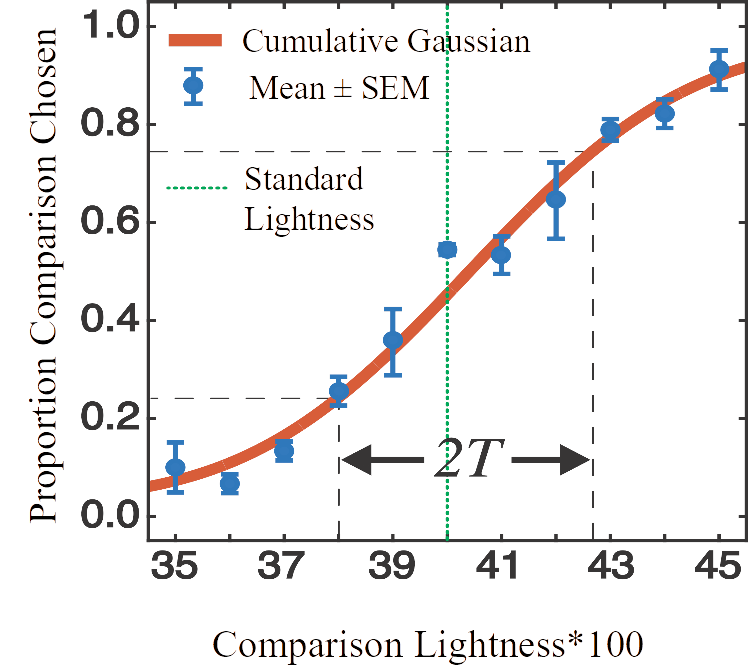
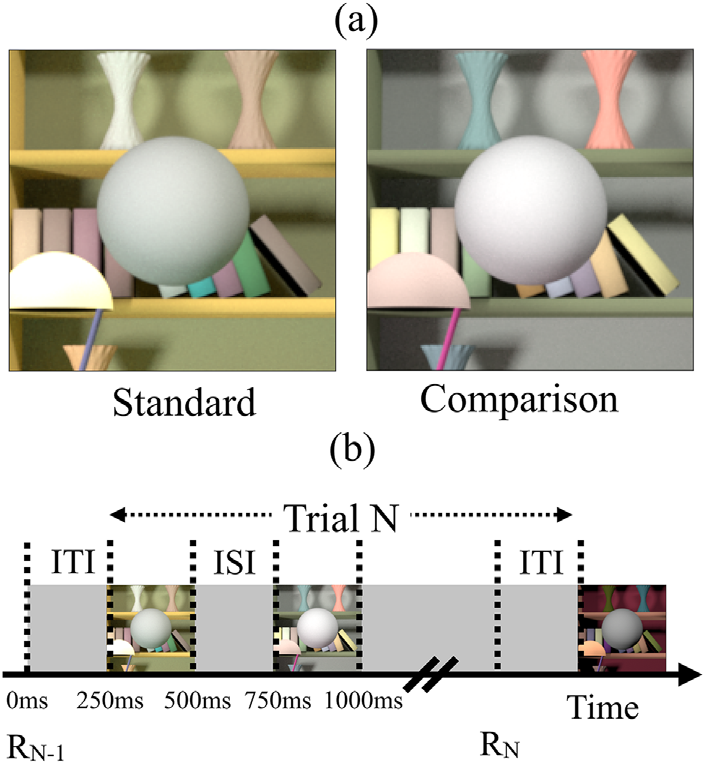


Figure 2. Psychophysical Task. (a) Sets of two naturalistic rendered images are displayed to a human subject. The subject reports which of the two images contain the spherical target object that is lighter. (b) The images are displayed consecutively for 0.25 seconds each. The inter-trial interval (ITI) and inter-stimulus interval (ISI) are both 0.25 seconds.

Figure 3. Psychometric Function. The mean of the proportion of times that the comparison image was chosen was plotted by the comparison lightness. For each of the 11 linearly spaced vales of the comparison lightness, the subjects reports of the lightness of 30 sets of images. Lightness discrimination thresholds are then chosen to be the 50%-76% difference of the comparison image lightness along a fitted cumulative gaussian.

The code to run the task, written in C#, is available on GitHub (https://github.com/DevinReynolds/ColorConstancyVR). Any scripts must be attached to objects inside of the scene to be run. Those scripts will be ran when the object they are attached to is loaded inside of the scene.

**Pilot Experiment**

To test the experiment, two pilot conditions are used: no variation and natural variation (Appendix Figure 1). These conditions have been previously studied with a monitor. In the case of no variation, the reflectance spectra of the background objects are held constant between images. In the case of natural Variation, the spectra of backgrounds objects are allowed to change in their overall shape (see: Image Rendering). The previous work concluded that the thresholds decrease as the amount of background variation increase, which predicts that the threshold of no variation should be lower than natural variation.

**Stimuli Presentation**

Images were presented to a subject on a Meta Quest 2 VR headset, that hasn’t been color calibrated. Images were be displayed consecutively. Each image was presented for 2.5 s, with 2.5 s between displaying each image, and 2.5 s between the subject’s decision and displaying the next image. Subject choice was forced and was collected once both images had been displayed and removed from the screen. The subject could take as much time as they wish before reporting their response. Subject response was collected using a keyboard.

**Data Collection**

For both conditions, one subject reported on the lightness of the target object 30 times per the 11 comparison object lightness values. Subject response was collected and saved as an XML file (Appendix Figure 2).

**Results**

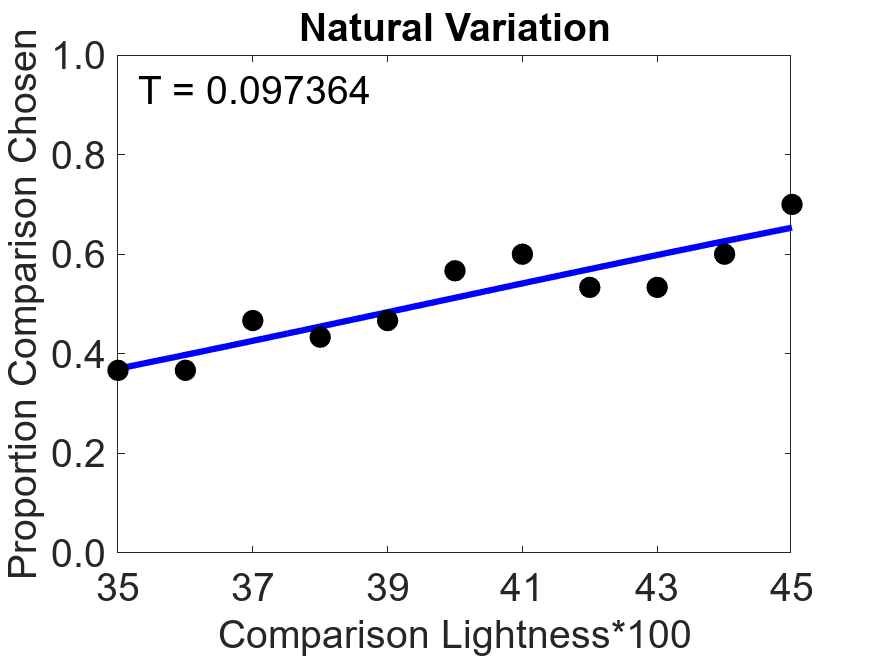
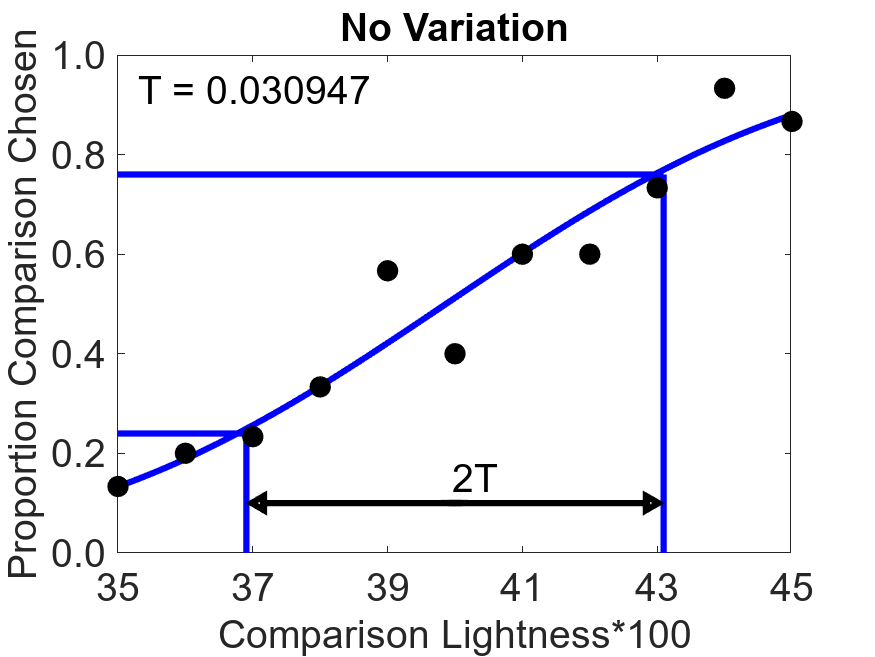
Lightness discrimination thresholds are greater in the case of natural variation compared to the case of no variation.

Figure 4. Psychometric Functions. For each of the 11 linearly spaced vales of the comparison lightness, the subject reports of the lightness of 30 sets of images. Lightness discrimination thresholds are then chosen to be the 50%-76% difference of the comparison image lightness along a fitted cumulative gaussian. (a) For no variation, the reflectance spectra of the background objects are constant between images. (b) For natural variation, the reflectance spectra of backgrounds objects are randomly sampled from the distribution of natural reflectances. This allows them to change in their color.

(a)

(b)

In previous monitor based experiments, lightness discrimination thresholds, under variation in background object reactance spectras, remain constant with relatively low amounts of variation and increase as the amount of variation increases. The VR based results are consistent with the monitor based.

**Conclusion/Future Work**

These results were found without color calibrating the VR headset while the monitor experiments results were found on a color calibrated monitor. As the headset has yet to be color calibrated, it’s not certain that what’s actually being displayed is the same as what was wanted to be displayed. This can lead to a difference between the thresholds of the VR set-up and the monitor set-up. Once the headset has been calibrated, thresholds can be measured and be compared to thresholds of the monitor based experiment. If the thresholds are the same, then it can be concluded that lightness discrimination thresholds do not depend on 3D information. If the thresholds are different, further extensive testing can be done to figure out what’s causing the difference.

**Acknowledgment**

Joel Oluyole, Jonathan Rinaldi, and Chibuike Okekeogbu help in the beginning in designing the VR set-up.

**References**

Brainard D. H. (2009). Color constancy. In Goldstein B. (Ed.), The Sage encyclopedia of perception (pp. 253–257). Los Angeles: SAGE Publications.

Cassin, B. & Solomon, S. (1990). Dictionary of Eye Terminology. Gainesville, Florida: Triad Publishing Company.

McCoun, J., & Reeves, L. (Eds.). (2010). Binocular vision : Development, depth perception and disorders : development, depth perception and disorders. Nova Science Publishers, Incorporated.

Singh, V., Cottaris, N. P., Heasly, B. S., Brainard, D. H., & Burge, J. (2018). Computational luminance constancy from naturalistic images. Journal of Vision

Singh, V., Burge, J., & Brainard, D. H. (2022). Equivalent noise characterization of human lightness constancy. *Journal of Vision*, *22*(5). https://doi.org/10.1167/jov.22.5.2

Appendix

Standard

Comparison

No Variation

Left

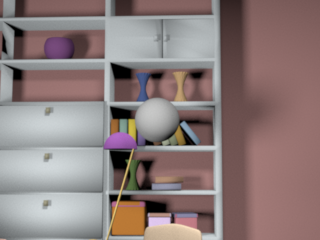
Right



Standard

Comparison

Natural Variation



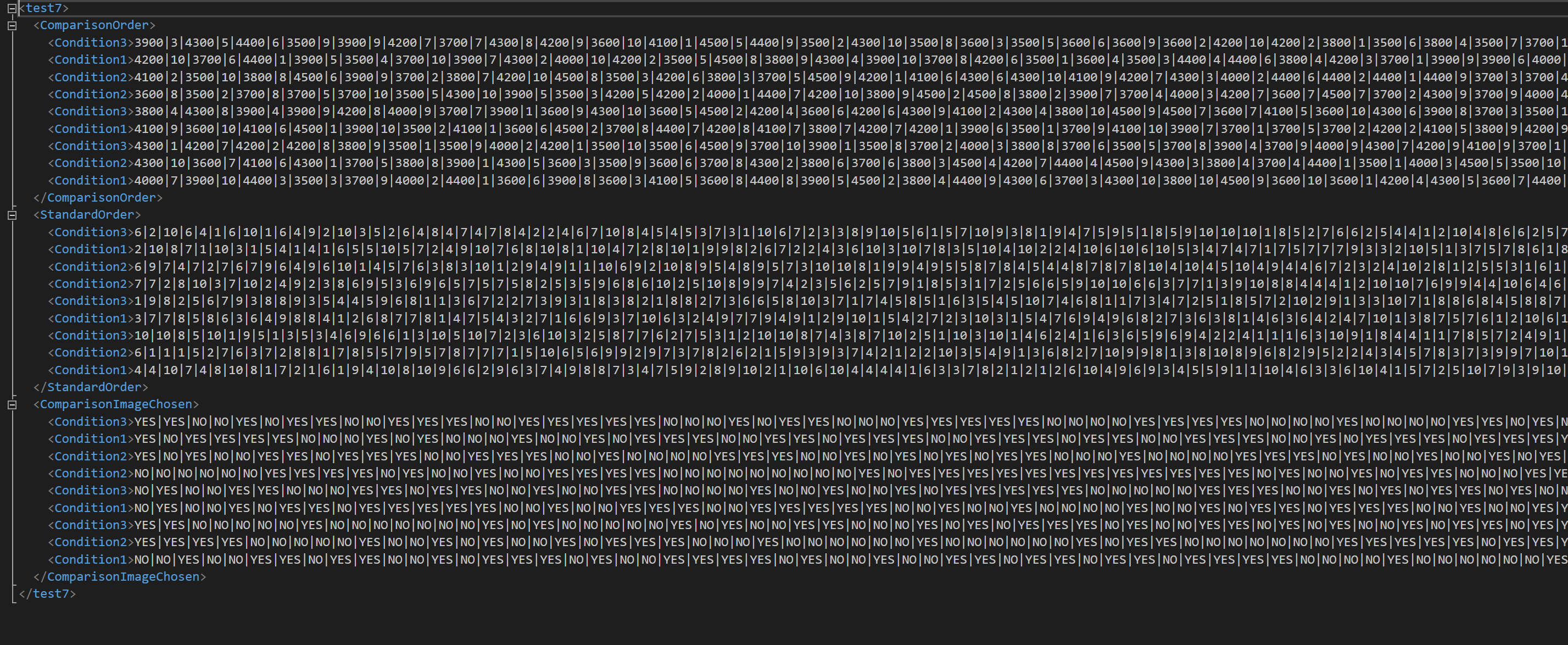
Left

Right

(a)

(b)

Appendix Figure 1. Sample stereoscopic pairs for two pilot conditions are used: no variation and natural variation. The left and right images are shows, however, when displayed on a VR headset these images are combined into a single by the human visual system. In the comparison pair, the target object can have one of 11 evenly spaced lightness values from 0.35 to 0.45. In the standard pair, the target object has a fixed lightness value of 0.40. (a) For no variation, the color of the backgrounds objects are fixed between the images. (b) For natural variation, the color of the background objects are allowed to change.

Appendix Figure 2. Sample XML file for a test experiment. In an XML file, data is organized into a structured format of elements. Each element has a name and can contain attributes and or child elements. Elements can also have values, which are the actual data being stored. The first time a session for a particular subject is ran, an XML file for said subject will be created. When the file is created, the StandardOrder, ComparisonOrder, and ComparisonImageChosen child nodes will be created. The StandardOrder and ComparisonOrder will be filled with a random order of conditions and random order of images displayed. ComparisonOrder will be populated as ‘lightness|image|lightness|image|…’, where lightness is one of the 11 values of the comparison image target object and image is the number of the image from the library generated. StandardOrder will be populated as ‘image|image|image|…’ as each standard image uses the same lightness values of 0.40 and doesn’t need to record which lightness values each time. The ComparisonImageChosen will be filled as sessions are completed with either ‘YES’ or ‘NO’ depending on whether or not the comparison image was chosen. The date and time a session was completed will be added on to the ComparisonImageChosen after the data.